tions were based in Asia (k=4), North America (k=3), Africa (k=3) and Europe (k=2). The target population of these publications were the general public (k=7), patients (k=4) or both (k=1). A slight majority of the 12 primary evidence publications (k=7) and a larger majority of the 5 secondary evidence publications (k=4) reported a SP bias on the results of the BG. Various parameters such as male gender, higher education and higher income levels were, in some instances, associated with higher WTP amounts. Other factors analysed are the population surveyed (patients vs. general population), and the location of the study. Association between these factors and the occurrence of starting point bias is examined and will be reported. **CONCLUSIONS:** There is evidence in the literature of a SP bias on the results of BGs, without however a full consensus on the matter. Further research is warranted in order to evaluate the conditions under which such bias appears.

PRM40

A LOOK AT PREVIOUS AND CURRENT METHODS USED TO COLLECT PATIENT-REPORTED OUTCOMES INFORMATION

Bergstrom F[1], Bolognese J[1], Tran D[2]
[1]*Covance Market Access Services, Inc., Gaithersburg, MD, USA,* [2]*Covance Market Access Services, Inc., San Diego, CA, USA*

OBJECTIVES: Patient reported outcomes (PROs) have become an important component of many clinical studies. The use of ePRO as a data collection method can alleviate the potential burden experienced by patients and/or sites. The purpose of this survey study was to capture current PRO data collection trends and summarize these findings side-by-side with results from a previous PRO data collection survey. **METHODS:** Industry professionals were invited to complete a web-based survey fielded in late 2011 and early 2012. This survey included questions on professional demographics, experience using PROs (and ePROs) by study type and experience with ePRO technologies. Responses were analyzed descriptively. **RESULTS:** To date, 54 industry professionals completed the 2011-2012 survey. Fifty nine percent of respondents were from pharmaceutical companies, biotech (26%), medical device (9%), and other (6%). While 49% of respondents in the 2010 survey had previous PRO study experience, 60% of respondents in the current survey had previous PRO experience. The proportion of respondents with prior ePRO experience, however, was similar across the two surveys (51% in 2010 and 54% in 2011-2012). Hand-held device (tablet, PDA) was the most common ePRO technology (71% in 2011, 64% in 2010), followed by interactive voice response (47% in 2011, 60% in 2010), and interactive web-response (29% in 2011, 51% in 2010). Among those with prior ePRO exposure in 2011 and 2010, respectively, 59% and 86% strongly agreed/agreed they would use ePRO in future studies. Among those who never used ePROs, 58% in 2011 and 50% in 2010 indicated they would likely use ePROs in future studies. **CONCLUSIONS:** Results from this survey suggest that ePRO use continues to gain moderate acceptance among industry professionals. These findings, however, were based on a limited sample size. Future surveys should be administered to allow future trends in ePRO use to be observed over time.

PRM41

TRANSLATION AND CULTURAL ADAPTATION OF THE LANGUAGE DEVELOPMENT SURVEY

Griffin A[1], Wild D[1], Rescorla L[2], Sunderaju R[3]
[1]*Oxford Outcomes, Oxford, Oxfordshire, UK,* [2]*Bryn Mawr College, Bryn Mawr, PA, USA,* [3]*University of Vermont, Burlington, VA, USA*

OBJECTIVES: The Language Development Survey (LDS) assesses children's word combinations and vocabulary and provides an accurate picture of a child's developing language when completed by a parent or guardian. The LDS contains a list of 310 basic words. The person completing the questionnaire is asked to circle those words their child says spontaneously. Translations already existed in over ten languages. Twelve further languages were translated, including eight Indic languages. A direct translation of some source words was not possible as some items are unknown in the target country, e.g. pretzel, spaghetti, Sesame St. Therefore, it was necessary to find a conceptually equivalent source word. **METHODS:** Two approaches were adopted: An initial translatability assessment was carried out to identify problematic words. Equivalent source words were suggested. Further problematic words were identified during the translation and cognitive-debriefing process. Alternatives were suggested either during the translation stages or by the parent/guardian during the cognitive debriefing interview. Final wording was agreed on through discussion with the lead in-country translator and instrument developers. **RESULTS:** A number of cultural adaptations were made. For all Indian languages, 'cracker' was translated as 'papadom' (a thin, crisp Indian cracker) and 'pizza' as translated as 'dosa' (a type of Indian pancake). This was decided before the initial translation step. 'Sesame St.' was replaced with 'Tom and Jerry'. In French 'saucisse' was suggested as an alternative for 'hot dog' during cognitive debriefing. **CONCLUSIONS:** When translating a patient-reported outcome (PRO) the aim is to produce a translation that is conceptually equivalent to the source text. In some cases, cultural adaptation is essential. Translation of the LDS is an excellent example of this methodology and the translations are now available for use in multi-national studies.

PRM42

PATIENT PREFERENCES FOR REMINDERS IN CLINICAL TRIALS: IMPROVING BOTH COMPLIANCE AND PATIENT EXPERIENCE

Ross J, Ross E
Almac Clinical Technologies, Souderton, PA, USA

OBJECTIVES: This session intends to identify patients' preferences for receiving reminders when participating in clinical trials. Results of a survey on patient experiences will be reported. Recommendations for developing reminder strategies will be provided with the intention of enhancing the patient's experience and compliance. **METHODS:** An internet survey was administered to patients (in December 2010) who participated in at least one clinical trial with patient diaries in the past two years. The survey focused on patients' perceived experiences and preferences with patient diaries/ePRO, and how patient diary methods could be improved. This session will focus on the questions relating to patient reminders–specifically relating to preferences for how reminders were sent/received, activities for which reminders were found most useful, frequency of reminders, etc. **RESULTS:** Reminder methods patients preferred most were email and text messages. Two areas patients preferred to be reminded about were any action required of them as part of their clinical trial participation and when they were required to record an electronic-diary entry. When patients were asked about how often they wanted to receive reminders, the most frequent responses included whenever there was new information related to their trial participation and once per day. Patients also provided suggestions for reminders in future trials. **CONCLUSIONS:** The results focus on what patients are telling industry what they do/do not want in terms of reminders; these perspectives should be accounted for to enhance the patient journey and compliance. If industry implements the reminder strategy wrong (for instance, annoying patients by reminding them too frequently), that may actually impact compliance negatively. Remind patients when necessary/not too often. Remind patients in ways they will be able to best receive/notice them. Appropriate use of reminders drives compliance and incorporating patient preferences will not only improve compliances rates, but will also enhance the patient's experience.

RESEARCH ON METHODS – Statistical Methods

PRM43

ROBUSTNESS OF CONFIDENCE INTERVALS FOR RARE EVENTS

Su Z[1], Mendelsohn A[1], Kim J[2], Gemmen E[3]
[1]*Outcome Sciences, a Quintiles Company, Cambridge, MA, USA,* [2]*Quintiles Global Consulting, London, Reading Berkshir, UK,* [3]*Quintiles Global Consulting, Rockville, MD, USA*

OBJECTIVES: Accurately estimating the upper bounds of confidence intervals for rare events such as hospitalization or death is an important activity in safety studies and outcomes research. Confidence intervals, however, for rare events are subject to considerable variation based upon the overall sample size and total number of observed events. This has led to a challenging convention that a minimum of 2 or 3 events are needed for computing meaningful confidence intervals. The objective of this study was to quantify the variation of the upper bound of confidence intervals for a binomial proportion in the setting of rare events. **METHODS:** Clopper-Pearson confidence intervals were constructed for sample sizes ranging from 50 to 1000, and numbers of events from 0 to 5. The robustness of the confidence interval was evaluated by calculating additional confidence intervals assuming: 1) one more observed event than in the original sample and, 2) that the proportion of events is equal to the upper bound of the confidence interval for the original sample. **RESULTS:** With sample sizes of 50, 100, 200, 500 and 1000, the upper bounds of the confidence intervals were 13.71%, 7.04%, 3.57%, 1.44% and 0.72%, respectively, with 2 observed events in the original sample; 16.55%, 8.52%, 4.32%, 1.74% and 0.87%, respectively, (3 observed events); and, 26.40%, 13.94%, 7.16%, 2.91% and 1.47%, respectively, when the proportion of events was equal to the upper bound of the confidence interval for the original sample with 2 events. Similar trends were seen when using other numbers of observed events. **CONCLUSIONS:** The upper bounds of confidence intervals for rare events vary greatly with sample sizes and the numbers of events observed when the sample size is small. A minimum of 500 subjects is optimal for constructing confidence intervals for rare events, even if 2 events or less are observed.

PRM44

PSEUDO-RANDOMIZATION IN RETROSPECTIVE ANALYSIS USING THE GENERALIZED MULTINOMIAL LOGIT FOR PROPENSITY SCORE GENERATION

Wilson SM, Mayne T
DaVita Clinical Research, Minneapolis, MN, USA

OBJECTIVES: To develop and test a three-way propensity score matching algorithm to provide pseudo-randomization of subjects into three groups to allow for comparable groups in a retrospective study. **METHODS:** Logistic regression using the generalized multinomial logit linking function was used to calculate estimates of the propensity score: the probability of having received three putatively interchangeable drugs from demographic (Race, Gender, Age) and comorbidities (Charlson Comorbidities Index) in a large, retrospective database. The most costly drug was used as the reference group, and the probability of each treatment group having received the reference drug was retained as the propensity score. In the initial analysis 23,912, 4,789, and 4,318 individuals were available in the three treatment regimens. Random subsets of 1/4 and 1/10 the original sample were constructed for the purpose assessing multi-group propensity score matching (PSM) effectiveness in constructing comparable groups via pseudo-randomization with varying starting sample sizes. PSM was conducted using calipers ranging from 8 digits to one digit of propensity score. Assessment of among-group differences before and after PSM were conducted using Chi-square tests for categorical variables and GLM analysis, with difference scores and their confidence intervals for continuous variables. **RESULTS:** For all sample sizes, prior to propensity score matching, significant differences existed among the three treatment regimens for all variables: gender, race, age and comorbidities. Following PSM there were 3381 matched triplets in the full sample. There were no significant differences among groups for gender, age or comorbidities; there were significant but tiny differences that remained for racial representation. In the smaller samples, 966 and 416 matched triplets were retained. There were no significant differences on any